DOCUMENT RESUME

ED 327 574 TM 016 021

AUTHOR	Reinhardt, Brian M.
TITLE	Factors Affecting Coefficient Alpha: A Mini Monte Carlo Study.
PUB DATE	15 Jan 91
NOTE	32p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (San Antonio, TX, January 26, 1991).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Computer Simulation; *Correlation; Difficulty Level; Estimation (Mathematics); Factor Analysis; *Item Response Theory; *Monte Carlo Methods; Scores; Test Items; *Test Reliability
IDENTIFIERS	*Alpha Coefficient; *Internal Consistency; Test Homogeneity; Variance (Statistical)

ABSTRACT

Factors affecting a lower-bound estimate of internal consistency reliability, Cronbach's coefficient alpha, are explored. Theoretically, coefficient alpha is an estimate of the correlation between two tests drawn at random from a pool of items like the items in the test under consideration. As a practical matter, coefficient alpha can be an index of internal consistency (i.e., an index of the degree to which item response scores correlate with total test scores). Cronbach's procedure is a method of calculating coefficient alpha. Three factors affecting coefficient alpha--total test variance, sum of item variances, and homogeneity of item difficulty--were examined using a mini Monte Carlo model. Small data sets generated by a computer program were used to make the discussion concrete. By considering bivariate correlation, multiple R correlation, beta-weight, and structure coefficients, total test variance was shown to account for the most variance of coefficient alpha, followed by the standard deviation of item difficulties, and the sum of item variances. Seven tables and three figures illustrate the discussion. (SLD)

bralpha.wp1 1/15/91

Factors Affecting Coefficient Alpha:

A Mini Monte Carlo Study

Brian M. Reinhardt

Texas A & M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, January 26, 1991.

ABSTRACT

The paper explores the factors that influence coefficient alpha as a lower bound estimate of score reliability. The purpose of the paper is to communicate the relative influence on alpha of various factors. Small data sets generated by a computer program are employed to make the discussion concrete and readily accessible.

Reliability is critical in detecting effects in substantive research. For example, if a dependent variable is measured such that it is perfectly unreliable, the effect size in the study will unavoidably be zero, and the results will not be statistically significant at any sample size, including an infinite one. As Locke, Spirduso and Silverman (1987, p. 28) note, "the correlation between scores from two tests cannot exceed the square root of the product for reliability in each test." Thus, if a researcher is correlating scores having a reliability of .9 with scores having a reliability of .6, the correlation cannot exceed .73. Prospectively, researchers must select measures that will allow detection of effects at the level desired; retrospectively, researchers must take reliability into account when interpreting findings.

Meier and Davis (1990), however, recently reported that published studies often do not adequately report reliability estimates so that results can be carefully evaluated. This has historically been the case with respect to published research. For example, Willson (1980) found that almost half the published studies he examined did not report reliability information. The same pattern seems to occur in dissertation research (LaGaccia, 1991).

The present paper explores the factors affecting a lower-bound estimate of internal consistency reliability, Cronbach's coefficient alpha. One aspect of this treatment involved use of a BASIC computer program (Thompson, 1990) that implements a small

scale Monte Carlo study for heuristic purposes.

What is coefficient alpha?

When an estimate of the reliability of scores on a test is needed and the parallel forms and test-retest approaches are impractical, researchers typically rely on internal consistency coefficients such as coefficient alpha (Feldt, Woodruff, & Salih, 1987). Theoretically, coefficient alpha is an estimate of the correlation expected between two tests drawn at random from a pool of items like the items in the test under consideration. Practically, coefficient alpha can be used as an index of internal consistency, i.e., an index of the degree to which item response scores (e.g., "0" or "1" in achievement testing) correlate with total test scores. Crocker and Algina (1986, p. 142) describe coefficient alpha as

> ...not a direct estimate of the [theoretical] reliability coefficient but rather an estimate of the lower bound of that coefficient....Alpha is the mean of all possible split-half coefficients that are calculated.

Alpha, therefore, can be interpreted as the lower bound estimate of the proportion of variance in the test scores explained by common factors underlying item performance.

Alpha is superior to the use of the split-half estimate of internal consistency, because for most tests of any length there are usually many splits, and the estimates associated with different splits for the same data may well yield contradictory

results. As Brownell (1933) pointed out long ago, for a test with k items, there are $(.5(k!))/[(.5(k))!]^2$ splits of the items. For example, for a test with six items, the number of splits (1,2,3 vs 4,5,6, 1,2,4 vs 3,5,6, etc.) equals:

$$\frac{.5 \times (6 \times 5 \times 4 \times 3 \times 2 \times 1)}{(3 \times 2 \times 1)^2} = \frac{.5 \times 720}{6^2} = \frac{360}{36} = 10$$

For a test with $k=2$ items there is one unique split. For a test with $k=4$ items there are three splits. For a test with six items, there are 10 splits, as indicated previously. For a test with 10 items there are 126 splits. Clearly the number of splits (and the likelihood of contradictory results across splits) escalates dramatically as items are added, even for relatively short tests.

Coefficient alpha has the same value no matter which method of computation is used. With this advantage, however, comes some cautions when interpreting alpha. Crocker and Algina (1986) warn that a relatively high value of coefficient alpha is not related to stability of the test scores over time, or to their equivalence to scores on one specific alternate form of the test, or to the unidemensionality of test items (i.e., performance on these items cannot necessarily be explained in terms of a single underlying factor).

This last limitation actually represents a common methodological error made by many researchers. A large alpha does not indicate that the test is unidimensional, and cannot be factor analyzed, notwithstanding the propensity of some researchers to formulate exactly this interpretation.

How is coefficient alpha calculated?

Three procedures used to calculate coefficient alpha determined from a single administration of a test are (a) Hoyt's analysis of variance, (b) Cronbach's alpha, and (c) Kuder-Richardson formula #20. All three methods yield identical results. Hoyt's method (Hoyt, 1941) is based on the analysis of variance, treating persons and items as sources of variation. Cronbach's alpha (Cronbach, 1951) uses item variance, total test variance, and the length of the test to compute an estimate of the internal consistency of items which are either dichotomously scored (i.e., scored either "0" or "1"), or which have a wide range of scoring weights. Cronbach's alpha is computed by the formula:

$$\alpha = \frac{k}{k-1} * [1 - \frac{\Sigma\sigma_i^2}{\sigma_T^2}] \qquad (\sigma_T^2 > 0,\ k > 1)$$

where k is the number of items on the test, $\Sigma\sigma_i^2$ is the sum of item variances, and σ_T^2 is the total test variance.

The Kuder-Richardson formulat #20 (Kuder & Richardson, 1937) is equivalent to Cronbach's alpha; however, KR 20 can only be used with dichotomously scored items. The KR 20 formula is identical to Cronbach's except the sum of item variances can be computed by using a simplified formula, $\sigma_i^2 = pq$, where p is the proportion of subjects correctly answering a given item (sometimes called item difficulty) and q (the proportion of subjects answering the item wrong) is 1 - p:

$$\alpha = \frac{k}{k-1} * [1 - \frac{\Sigma pq}{\sigma_T^2}]$$

For a complete algebraic proof of the formula, $\sigma_i^2 = pq$, for

dictomously scored items, see Crocker and Algina (1986, pp. 90-92).

For dichotomously scored items, item difficulty (p) can only range between 0 and 1 inclusive, since for a given item j no fewer than a proportion of 0 (or 0%) of the subjects can get the item right and no more than a proportion of 1 (or 100%) of the subjects can get the item right. Though it is perhaps counter-intuitive, larger p-values indicate easier items, and smaller p-values describe more difficult items. KR 20 will be used to calculate coefficient alpha in this study because all of the test items used are dichotomously scored.

What factors affect coefficient alpha?

Both the characteristics of the person sample selected and the characteristics of the test items can affect coefficient alpha. For example, if a group of examinees is homogeneous with respect to knowledge about physics, that is they have similar competencies in this area, then one would expect them to score about the same on a test of physics concepts. In this example, the variability of total test scores would be small because most examinees would have a similar score. This would result in dividing each item variance by a smaller number, yielding a larger result that is then subtracted from one, and thereby yielding a smaller estimated reliability, as can be seen by examining the KR 20 formula. Thus, though it is not widely recognized by researchers, the nature (and even the size) of the sample of subjects can impact the estimated classical reliability coefficient.

In considering the characteristics of the test that may affect

coefficient alpha, it is important to note that a test is not reliable or unreliable; rather, "reliability is a property of the scores on a test for a particular group of examinees" (Crocker & Algina, 1986, p. 144, emphasis added). Depending on the types of items added, increasing the length of the test usually increases coefficient alpha, that is if (and only if) the items are of equal or better quality than the other items on the test. The more items that are selected from the "pool of items," the less likely that the item sample will be biased, or will underestimate the theoretical coefficient alpha.

In the computational equation for coefficient alpha, the k/(k-1) term corrects for this bias; when the number of items on the test, k, increases, then this correction term gets smaller and smaller. For example, for a 2-item test, the correction multiplier would be (2/(2-1)) or 2; with a 10-item test, k equals 1.11; with a 100-item test, k equals 1.0101. There is a point when adding items will help less and less to increase coefficient alpha as regards this correction aspect of the formula.

Three of the factors that affect coefficient alpha, total test variance, sum of item variances, and homogeneity of item difficulty, p, will be examined in the present study using a mini Monte Carlo model. The Monte Carlo method is a technique for obtaining an approximate solution to certain mathematical and logical problems; it characteristically involves randomly sampling from some specified and known universe. The process is usually done on a computer, as it was in this study.

Characteristics of the modelled items sets are easily manipulated using this method; the method allows use of a wide variety of item set types to help increase the generalizability of the results. For the present study, *four population target matrices or models* were used to manipulate the combination of (a) total test variance, (b) sum of item variances, and (c) homogeneity/heterogeneity of item difficulty for a hypothetical 7-item, dichotomously scored test with 10 examinees. The purpose of this Monte Carlo modeling was <u>to investigate the relative effects of total test variance, item variance, and homogeneity of item difficulty with respect to their influences on reliability</u>.

Each population target model creates the conditions necessary to maximize, minimize or moderate each of the three factors that affect coefficient alpha. As part of the random sampling process, a probability is specified to influence the degree to which random samples will reflect each parameter in the population. For example, if the assigned probability is .2 for item 2 of person 2, then when a random number is generated to determine whether the score for this item in a given random sample will be a "0" or a "1," on the average 20 percent of the time this item will be scored 1 and 80 percent of the time this item will be scored a 0. Table 1 presents the four population models used in the study.

Insert Table 1 about here

As can be determined by consulting Table 1, the four models create various patterns in the total test variance, sum of item

variances, and homogeneity/heterogeneity of item difficulty. Model 1 produces maximum total score variance, maximum item variance, and homogeneous item difficulty. Model 2 generates minimum total score variance, maximum item variance, and homogeneous item difficulty. Model 3 produces moderate total score variance, moderate sums of the item variances, and heterogeneous item difficulty. Model 4 generates minimum total variance, moderate sums of items variances, and heterogeneous item difficulty.

How can total test variance be maximized?

On a 7-item, dichotomously scored test with 10 examinees, the total test score for each examinee ranges from 0 to 7. Figure 1 uses histograms to illustrate various total test score combinations and their corresponding total test variance. When everyone gets a score of "1" for each item, each examinee's total test score is 7. The item difficulty, p, for each item in this example is 1 and the total test variance, σ_T^2, is 0, because there is no "spreadoutness" in the scores. Similarly, when everyone gets a score of 0 for each item, each examinee's total score is 0, item difficulty, p, for each item is 0, and once again, total test variance, σ_T^2, is 0, because there is no "spreadoutness" in the scores.

Insert Figure 1 about here

The third histogram in Figure 1 shows that when each examinee earns a different score (i.e., all of the possible total test scores are represented), then the total test variance is approximately 3. The maximum total test variance possible on a 7-

item, dichtomously scored test with 10 examinees, however, is much higher than 3. Total test variance is maximized when the sum of squares is maximized. This occurs when the total test scores are "spreadout" the most, or in other words, <u>when half of the examinees earn the lowest possible total score (here it's 0) and half earn the highest possible total score (here it's 7)</u>. When half of the examinees get all of the items "correct" and half "miss" all of the items, then the total test variance is maximized as 12.25. The range of the total test variance on a 7-item dichotomously-scored test, then, is from 0 to 12.25.

By examining the total test scores of the <u>expected or population</u> model in Table 2b, one can determine that Model 1 will maximize the total test variance, because half of the examinees in the population earn total test scores of 0 and half earn total test scores of 7. Table 3b shows that Model 2 produces a minimal total test variance, because half of the examinees earn a total test score of 3 and the other half earn a total test score of 4. The minimal "spreadoutness" yeilds a total test variance of .25 in the actual population from which the 10 random samples for Model 2 were drawn.

Insert Tables 2 and 3 about here

Table 4b demonstrates that Model 3 yeilds a moderate total test variance of 3.36, because half of the examinees in the population earn a total test score of 0, one-forth earn a total test score of 3, and one-forth earn a total test score of 4. Table

5b shows that Model 4, like Model 2, minimizes the total test variance. In the population from which the 10 data sets were randomly drawn for Model 4, all of the examinees earn a total test score of 2. With no "spreadoutness" in the scores, the total test variance is 0.

Insert Tables 4 and 5 about here

How can the sum of item variances be maximized?

A method similar to the one just described to find maximum total test variance can be used to determine the maximum sum of item variances. Consider the scores on a given item j to be analogous to the scores on the total test. To maximize the variance of a single item j, one must maximize the "spreadoutness" of scores which is measured by using the the sum of squares. From the previous discussion, it was shown that if half of the scores are at one extreme of the range and half are at the other extreme, then the sum of squares, and therefore, the variance, is maximized. Since the possible scores for each examinee on a given item j are either "0" or "1", then to maximize the item's variance, half of the examinees would earn a 0 and half would earn a 1.

Figure 2 shows that items with moderate difficulty (i.e., p = .5) maximize item variance and items that are either easier (e.g., p = 1) <u>or</u> harder (e.g., p = 0) tend to minimize item variance. Item variances can range from 0 to .25; therefore, the maximum sum of item variances for a 7-item, dichotomously scored test is (.25 times 7) or 1.75.

Insert Figure 2 about here

Table 2b shows that the expected or population model for Model 1 yeilds the maximum sum of item variances (1.75) because half of the examinees on each item earn a 0 and half earn a 1. Similarly, Table 3b shows that Model 2 maximizes the sum of item variances by the same method. Table 4b indicates that Model 3 produces a moderate sum of item variances ($\Sigma\sigma_i^2 = .96$), because on half of the items everyone earned a 0 and on the other half, half of the examinees earned a 0 and half earned a 1. A similar procedure was used in Model 4 (Table 5b) to yield a moderate sum of item variances ($\Sigma\sigma_i^2 = 1.0$).

What does homogeneous p and heterogeneous p mean?

Homogeneous p, or difficulty level, means that all the items have the same or similar difficulty level. This is equivalent to having the standard deviation of p be either zero or be very small. Models 1 and 2 both illustrate homogeneous p, because for each item, p equals .5. This means that half of the examinees earn a 0 on each item and half earn a 1.

Item difficulty, p, is heterogeneous when the items vary in difficulty, and thus, the standard deviation of *p* would be large. Both Models 3 and 4 illustrate heterogeneous p because half of the items have a p-value of .5 and the other half have a p-value of 0.

How do total test variance, sum of item variances, and homogeneity/heterogeneity of item difficulty affect alpha?

Using the mini Monte Carlo method, 10 random samples of the population models were generated to examine the affects on

coefficient alpha of manipulating the three target factors: (a) total test variance, (b) sum of item variance, and (b) homogeneity/ heterogeneity of item difficulty. Table 6 displays the results of the study, listing the mean of the alpha coefficients generated by the 10 random samples for each population model, along with the standard deviation, and minimum and maximum values for the sampled alphas. The table also lists the mean, standard deviation, and minimum and maximum values for the total test variances (TESTVAR), the sums of the item variances (IVARSUM), and the standard deviation of the item difficulties (ITEMPSD).

Insert Table 6 about here

Model 1 (maximum TESTVAR, maximum IVARSUM, small ITEMPSD) tended to yeild the highest alpha coefficients. The mean of the 10 alphas for Model 1 was .94, with a very small standard deviation of .026, indicating that most of the alpha coefficients clustered close to .94. Model 2 (moderate TESTVAR, moderate IVARSUM, small ITEMPSD) produced the second highest alpha coefficients ($\bar{\alpha}$ = .656, σ = .138). Model 2 (minimum, maximum, homogeneous) and Model 4 (minimum, moderate, heterogeneous) yielded the lowest coefficient alpha levels. In fact, both of these mean alpha values were negative (Model 2, $\bar{\alpha}$ = -.146, SD = .377; Model 4, $\bar{\alpha}$ = -.232, SD = .509).

These last two results make the point that alpha can be negative. Indeed, the population model reported in Table 3b yields a value of -7.0. These results make the point that alpha is a

lower bound on a true reliability estimate, since the coefficient does not behave as a correlation coefficient would.

From these results several tentative conclusions can be drawn. First, moderate to maximum total test variance is important to maximize coefficient alpha. It appears that total test variance accounts for much of the variance that explains coefficient alpha. Second, total test variance appears to have more affect on alpha that the sum of item variances. Third, the homogeneity/heterogeneity of item difficulty, p, seems to have minimal effect on alpha. To further define these tentative conclusions, the relationship of total test variance, sum of item variances, and item difficulty must be considered.

<u>**How can the relationship of total test variance, sum of item variances, and item difficulty be investigated?**</u>

To determine the portion of variance of alpha explained by each of the factors alone across the 40 sets of results (4 models x 10 random samples of results for each), the bivariate correlations between alpha and each of the factors separately can be computed. If Y is alpha, X_3 is total test variance, X_2 is sum of item variances, and X_1 is the standard deviation of item difficulties, then for the data in the present study r_{yx3} is .772, r_{yx2} is .083, and r_{yx1} is -.317. By squaring each of these bivariate correlation coefficients, one can determine the percent of variance that each uniquely contributes to alpha.

The factor, total test variance, explains 60% of the variance of alpha (r^2 = .5959); the factor, sum of item variances, explains less than 1% of the variance of alpha (r^2 = .0069); and the

standard deviation of item difficulties explains 10% of the variance of alpha ($r^2 = .10$). From this brief analysis, it appears that total test variance has the greatest affect on alpha, followed by the standard deviation of item difficulties, and finally, the sum of item variances.

A multiple regression analysis can assist in investigating the collective and separate contributions of the three independent variables, total test variance, sum of item variances, and the standard deviation of item difficulties, to the variation of the dependent variable, coefficient alpha (Howell, 1982, pp. 414-419). The multiple R correlation coefficient expresses the magnitude of the relation between the best possible combination of all independent variables (X_1, X_2, X_3) and the dependent variable (Y) (Kerlinger, 1979, pp. 171-172). R^2, which is similar to r^2, is the proportion of variance of Y accounted for by the regression combination of all the independent variables, excluding double counting of any area of Y jointly predicted by two or more of the predictors. Since in the present example R^2 was .73298, then 73% of the variance of alpha is explained by the combination of X_1 and X_2 and X_3.

The regression equation or prediction equation from the data in this study is:

$$Y' = .189984\, X_3 - 1.067318\, X_2 + .57629\, X_1 + 1.092742$$

where Y' is predicted alpha, X_3 is total test variance, X_2 is sum of item variances, and X_1 is the standard deviation of item difficulties. These results seem to suggest the incorrect

conclusion that the sum of the item variances is the best predictor of alpha.

But the B-weight coefficients for each of the independent variables cannot be compared because their units do not have the same scales of measurement. To alleviate this problem, the regression equation can be written in Z-score form (Howell, 1982, pp. 419-420). By converting the B-weights to ß-weights by using the formula $\beta = b\ (SD_x/SD_y)$, the effects of the different measurement scales are washed out, and the ß-weight coefficients can be compared. The new regression equation in Z-score form is:

$$Z'_Y = 1.014169\ Z_3 - .395216\ Z_2 + .053201\ Z_1 + 0$$

The unthoughtful researcher may wish to discuss the relative influence each independent variable has on alpha by considering only the ß-weight coefficients; however, Thompson and Borrello (1985) point out that especially with small to moderate sample sizes, ß-weight coefficients will fluctuate. This fluctuation is due to the back and forth distribution of shared variance among independent variables. In any case, these weights are influenced by the correlation among the predictor variables, and should not be the sole basis for interpreting regression results.

When predictor variables are correlated with each other, as most are, then "collinearity" or "multicollinearity" exists. Kerlinger (1979, p. 165) notes that "independent variables [in educational psychology research]...are almost always correlated, often substantially so." In this study, the independent variables are correlated with each other, and thus the ß-weight coefficients

should not be the only indices to use when interpreting relative influence on alpha of these three factors. Pedhazur (1982, p. 246) reiterates this point by stating that "the presence of high multicollinearity poses serious threats to the interpretation of regression coefficients [ß-weights] as indices of effects."

Thompson and Borello (1985, p. 208) note that "structure coefficients are not suppressed or inflated by collinearity " These coefficients, unlike ß-weight coefficients, can be used to describe the dependent variable in terms of its relationships with the independent variables. The structure coefficient for a predictor variable is equal to the bivariate correlation between the predictor variable and the dependent variable divided by the multiple R correlation coefficient:

$$\text{structure coefficient of } X_i = r_{yxi}/R$$

The structure coefficients for each of the independent variables in this study were .90172 for X_3, .09695 for X_2, and -.37027 for X_1.

Table 7 presents the rank orders of the influences on the 40 alphas for each factor using ß-weight coefficients and structure coefficients. Both coefficients suggest that total test variance is the most important influence on alpha. The structure coefficients indicate that the standard deviation of item difficulties ranked second in influence and the sum of item variances ranked third.

Insert Table 7 about here

By using the three predictor variables of total test variance,

sum of item variances, and the standard deviation of item difficulties, coefficient alpha was reasonably accurately predicted (R^2 = 73.3%) for these 40 sets of data. Figure 3 illustrates the relatively close fit between the 40 observed and predicted alpha values.

Insert Figure 3 about here

Summary

Using data generated by a mini Monte Carlo method BASIC computer program, the present study examined three factors (total test variance, sum of item variances, and the standard deviation of item difficulties) that affect coefficient alpha. By considering bivariate correlation, multiple R correlation, β-weight, and structure coefficients, total test variance was shown to account for the most variance of coefficient alpha, followed by the standard deviation of item difficulties, and the sum of item variances.

There are two reasons to suspect that these results will generalize. The first reason is <u>empirical</u>. The four population models represent a wide range of possible results that can occur in practice, suggesting that the ballpark of possibilities was covered in the mini Monte Carlo study. The second reason is <u>theoretical</u>. As can be surmised by examining the formula for KR 20, because total test variance can range widely in numerical values especially as test length increases, it should be expected that total test variance can dominate alpha. Thus, the results and the expository

treatment here suggest that researchers and practitioners should generally focus on total score variance when they are trying to maximize alpha.

References

Brownell, W. A. (1933). On the accuracy with which reliability may be measured by correlating test halves. Journal of Experimental Education, 1, 204-215.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Chicago: Holt, Rinehart and Winston.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of test. Psychometrika, 16, 297-334.

Feldt, L. S., Woodruff, D. J., Salih, F. A. (1987). Statistical inference for coefficient alpha. Applied Psychological Measurement, 11, 93-103.

Howell, D. C. (1982). Statistical methods for psychology. Boston, MA: Duxbury Press.

Hoyt, C. (1941). Test reliability estimated by analysis of variance. Psychometrika, 6, 153-160.

Kerlinger, F. N. (1979). Behavioral research: A conceptual approach. New York: Holt, Rinehart and Winston.

Kuder, G. F., & Richardson, M. W. (1937). The theory of estimation of test reliability. Psychometrika, 2, 151-160.

LaGaccia, S. (1991). Methodology choices in a cohort of education dissertations. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 149-158). Greenwich, CT: JAI Press.

Locke, L. F., Spirduso, W. W., & Silverman, S. J. (1987). Proposals that work: A guide for planning dissertations and grant proposals (2nd ed.). Newbury Park, CA: SAGE.

Mehrens, W. A., & Lehemann, I. J. (1984). Measurement and evaluation in education and psychology (3rd ed.). New York: Holt, Rinehart and Winston.

Meier, S. T., & Davis, S. R. (1990). Trends in reporting psychometric properties of scales used in counseling psychology research. Journal of Counseling Psychology, 37, 113-115.

Pedhazur, E. J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed.). New York: Holt, Rinehart and Winston.

Thompson, B. (1990). ALPHMONT: A computer program that simulates influences on coefficient alpha. College Station, TX: Texas A & M University.

Thompson, B., & Borrello, G. M. (1985). The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Willson, V.L. (1980). Statistical techniques in AERJ articles: 1969 to 1978. Educational Researcher, 9(6), 5-10.

Table 1

Four Models that Manipulate Factors Affecting Coefficient Alpha

Factor	Model 1	Model 2	Model 3	Model 4
Total Test Variance	MAX	MIN	MOD	MIN
Sum of Item Variances	MAX	MAX	MOD	MOD
Homogeneity/ Heterogeneity of p	HOM	HOM	HET	HET

Table 2

Probability Target Matrix (a), Population Model Matrix (b), and One of the 10 Randomly Sampled Matrices (c) for Model 1

a. MAX TEST VAR (G), MAX ITEM VAR (B), HOMOGENEOUS p (G)
PROBABILITY TARGET MATRIX

1	.1	.1	.1	.1	.1	.1	.1
2	.1	.1	.1	.1	.1	.1	.1
3	.1	.1	.1	.1	.1	.1	.1
4	.1	.1	.1	.1	.1	.1	.1
5	.1	.1	.1	.1	.1	.1	.1
6	.9	.9	.9	.9	.9	.9	.9
7	.9	.9	.9	.9	.9	.9	.9
8	.9	.9	.9	.9	.9	.9	.9
9	.9	.9	.9	.9	.9	.9	.9
10	.9	.9	.9	.9	.9	.9	.9

b. MAX TEST VAR (G), MAX ITEM VAR (B), HOMOGENEOUS p (G)
Known Results for the Population

n	1	2	3	4	5	6	7	TOTAL
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	1	1	1	1	1	1	1	7
7	1	1	1	1	1	1	1	7
8	1	1	1	1	1	1	1	7
9	1	1	1	1	1	1	1	7
10	1	1	1	1	1	1	1	7
p	.5	.5	.5	.5	.5	.5	.5	
var_i	.25	.25	.25	.25	.25	.25	.25	12.25

ALPHA = 1.166667 TIMES (1 - (1.75 / 12.25)) = 1

c. MAX TEST VAR (G), MAX ITEM VAR (B), HOMOGENEOUS p (G)

n	1	2	3	4	5	6	7	TOTAL
1	0	0	0	0	0	1	0	1
2	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	1
4	0	0	0	0	0	0	0	0
5	0	0	0	1	0	0	0	1
6	1	1	1	1	1	1	0	6
7	1	1	1	1	1	1	1	7
8	1	1	1	1	1	1	1	7
9	1	1	1	1	1	1	1	7
10	1	1	1	1	1	1	1	7
p	.6	.5	.5	.6	.5	.6	.4	
var_i	.24	.25	.25	.24	.25	.24	.24	9.809999

ALPHA = 1.166667 TIMES (1 - (1.71 / 9.809999)) = .9633028

Table 3

Probability Target Matrix (a), Population Model Matrix (b), and One of the 10 Randomly Sampled Matrices (c) for Model 2

a. MIN TEST VAR (B), MAX ITEM VAR (B), HOMOGENEOUS p (G)
PROBABILITY TARGET MATRIX

1	.6	.1	.6	.1	.6	.1	.6
2	.2	.8	.2	.8	.2	.8	.2
3	.8	.2	.8	.2	.8	.2	.8
4	.2	.8	.2	.8	.2	.8	.2
5	.8	.2	.8	.2	.8	.2	.8
6	.2	.8	.2	.8	.2	.8	.2
7	.8	.2	.8	.2	.8	.2	.8
8	.2	.8	.2	.8	.2	.8	.2
9	.8	.2	.8	.2	.8	.2	.8
10	.4	.9	.4	.9	.4	.9	.4

b. MIN TEST VAR (B), MAX ITEM VAR (B), HOMOGENEOUS p (G)
Known Results for the Population

n	1	2	3	4	5	6	7	TOTAL
1	1	0	1	0	1	0	1	4
2	0	1	0	1	0	1	0	3
3	1	0	1	0	1	0	1	4
4	0	1	0	1	0	1	0	3
5	1	0	1	0	1	0	1	4
6	0	1	0	1	0	1	0	3
7	1	0	1	0	1	0	1	4
8	0	1	0	1	0	1	0	3
9	1	0	1	0	1	0	1	4
10	0	1	0	1	0	1	0	3
p	.5	.5	.5	.5	.5	.5	.5	
var_i	.25	.25	.25	.25	.25	.25	.25	.25

ALPHA = 1.166667 TIMES (1 - (1.75 / .25)) = -7

c. MIN TEST VAR (B), MAX ITEM VAR (B), HOMOGENEOUS p (G)

n	1	2	3	4	5	6	7	TOTAL
1	1	0	1	0	1	0	1	4
2	1	1	0	1	0	1	0	4
3	1	1	1	0	1	1	1	6
4	1	1	0	1	0	1	0	4
5	1	0	1	0	0	0	1	3
6	0	1	0	1	0	1	0	3
7	1	0	1	1	1	1	1	6
8	0	0	1	0	0	1	0	2
9	1	0	1	1	1	1	1	6
10	0	1	0	1	0	1	1	4
p	.7	.5	.6	.6	.4	.8	.6	
var_i	.21	.25	.24	.24	.24	.16	.24	1.76

ALPHA = 1.166667 TIMES (1 - (1.58 / 1.76)) = .1193182

Table 4

Probability Target Matrix (a), Population Model Matrix (b), and One of the 10 Randomly Sampled Matrices (c) for Model 3

a. MOD TEST VAR (M), MOD ITEM VAR (M), HETEROGENEOUS p (B)
PROBABILITY TARGET MATRIX

1	.1	.1	.1	.1	[illegible]	.1	.1
2	.1	.1	.1	.1	[illegible]	.1	.1
3	.1	.1	.1	.1	[illegible]	.1	.1
4	.1	.1	.1	.1	.1	.1	.1
5	.1	.1	.1	.1	.1	.1	.1
6	.1	.1	.1	.9	.9	.9	.9
7	.1	.1	.1	.1	.9	.9	.9
8	.1	.1	.1	.9	.9	.9	.9
9	.1	.1	.1	.1	.9	.9	.9
10	.1	.1	.1	.9	.9	.9	.9

b. MOD TEST VAR (M), MOD ITEM VAR (M), HETEROGENEOUS p (B)
Known Results for the Population

n	1	2	3	4	5	6	7	TOTAL
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	1	1	1	1	4
7	0	0	0	0	1	1	1	3
8	0	0	0	1	1	1	1	4
9	0	0	0	0	1	1	1	3
10	0	0	0	1	1	1	1	4
p	0	0	0	.3	.5	.5	.5	
var_i	0	0	0	.21	.25	.25	.25	3.36

ALPHA = 1.166667 TIMES (1 - (.9600001 / 3.36)) = .8333333

c. MOD TEST VAR (M), MOD ITEM VAR (M), HETEROGENEOUS p (B)

n	1	2	3	4	5	6	7	TOTAL
1	0	0	0	0	1	1	0	2
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	1
5	0	0	0	0	0	0	0	0
6	0	0	0	1	1	0	1	3
7	0	0	1	0	1	1	1	4
8	0	0	0	1	1	1	1	4
9	0	0	0	0	0	1	0	1
10	1	0	0	1	1	1	1	5
p	.2	0	[illegible]	.3	.5	.5	.4	
var_i	.16	0	.09	.21	.25	.25	.24	3.2

ALPHA = 1.166667 TIMES (1 - (1.2 / 3.2)) = .7291666

Table 5

Probability Target Matrix (a), Population Model Matrix (b), and One of the 10 Randomly Sampled Matrices (c) for Model 4

a. MIN TEST VAR (B), MOD ITEM VAR (M), HETEROGENEOUS p (B)
PROBABILITY TARGET MATRIX

1	.1	.1	.1	.2	.7	.2	.7
2	.2	.2	.2	.2	.8	.2	.8
3	.2	.2	.2	.2	.8	.2	.8
4	.2	.2	.2	.2	.8	.2	.8
5	.2	.2	.2	.2	.8	.2	.8
6	.2	.2	.2	.8	.2	.8	.2
7	.2	.2	.2	.8	.2	.8	.2
8	.2	.2	.2	.8	.2	.8	.2
9	.2	.2	.2	.8	.2	.8	.2
10	.3	.3	.3	.8	.3	.8	.3

b. MIN TEST VAR (B), MOD ITEM VAR (M), HETEROGENEOUS p (B)
Known Results for the Population

n	1	2	3	4	5	6	7	TOTAL
1	0	0	0	0	1	0	1	2
2	0	0	0	0	1	0	1	2
3	0	0	0	0	1	0	1	2
4	0	0	0	0	1	0	1	2
5	0	0	0	0	1	0	1	2
6	0	0	0	1	0	1	0	2
7	0	0	0	1	0	1	0	2
8	0	0	0	1	0	1	0	2
9	0	0	0	1	0	1	0	2
10	0	0	0	1	0	1	0	2
p	0	0	0	.5	.5	.5	.5	
var_i	0	0	0	.25	.25	.25	.25	0

ALPHA = 1.166667 TIMES (1 - (1 / 0)) = -999

c. MIN TEST VAR (B), MOD ITEM VAR (M), HETEROGENEOUS p (B)

n	1	2	3	4	5	6	7	TOTAL
1	0	0	0	1	1	0	1	3
2	0	0	0	0	1	0	1	2
3	0	0	0	0	1	0	1	2
4	0	0	0	0	1	1	0	2
5	0	0	0	1	0	1	1	3
6	0	0	1	1	0	1	0	3
7	0	0	0	0	0	1	0	1
8	1	0	0	1	0	1	0	3
9	0	0	0	0	0	1	0	1
10	1	0	1	1	0	1	0	4
p	.2	0	.2	.5	.4	.7	.4	
var_i	.16	0	.16	.25	.24	.21	.24	.84

ALPHA = 1.166667 TIMES (1 - (1.26 / .84)) = -.5833333

Table 6

Results from the Four Models Using the Monte Carlo Method

VARIABLE	MEAN	STD DEV	MINIMUM	MAXIMUM	N
Model 1 (MAXIMUM, MAXIMUM, HOMOGENEOUS)					
ALPHA	.940	.026	.906	.984	10
TESTVAR	8.884	1.134	7.610	11.050	10
IVARSUM	1.702	.028	1.660	1.740	10
ITEMPSD	.074	.021	.035	.105	10
Model 2 (MINIMUM, MAXIMUM, HOMOGENEOUS)					
ALPHA	-.146	.377	-.718	.537	10
TESTVAR	1.564	.585	1.040	3.040	10
IVARSUM	1.606	.053	1.520	1.680	10
ITEMPSD	.134	.026	.099	.175	10
Model 3 (MODERATE, MODERATE, HETEROGENEOUS)					
ALPHA	.656	.138	.292	.778	10
TESTVAR	2.876	.475	1.600	3.200	10
IVARSUM	1.210	.083	1.030	1.320	10
ITEMPSD	.187	.024	.139	.225	10
Model 4 (MINIMUM, MODERATE, HETEROGENEOUS)					
ALPHA	-.232	.509	-.833	.539	10
TESTVAR	1.295	.536	.760	2.250	10
IVARSUM	1.353	.178	1.100	1.580	10
ITEMPSD	.186	.043	.120	.251	10

Table 7

Rank of Influence of Factors Affecting Coefficient Alpha

Coefficient Used	First	Second	Third
β-weight	X_3[a]	X_2[b]	X_1[c]
Structured	X_3	X_1	X_2

[a] X_2 = IVARSUM
[b] X_1 = ITEMPSD
[c] X_3 = TESTVAR

Figure 1
Histograms illustrating total test variance with a 7-item, dichotomously scored test with 10 examinees

```
               X
               X
               X
               X
               X
               X
               X
               X
               X
               X
_______________X_
0 1 2 3 4 5 6 7
```

Total Test Score
For each item, p = 1
$\sigma_T^2 = 0$

```
X
X
X
X
X
X
X
X
X
X
X_______________
0 1 2 3 4 5 6 7
```

Total Test Score
For each item, p = 0
$\sigma_T^2 = 0$

```
      X X
X X X X X X X X
_______________
0 1 2 3 4 5 6 7
```

Total Test Score
$\sigma_T^2 = 3.21$

```
X             X
X             X
X             X
X             X
X_____________X_
0 1 2 3 4 5 6 7
```

Total Test Score
$\sigma_T^2 = 12.25$

Figure 2

The 11 Possible Variances for a Dichotomously-Scored Item Completed by 10 Examinees

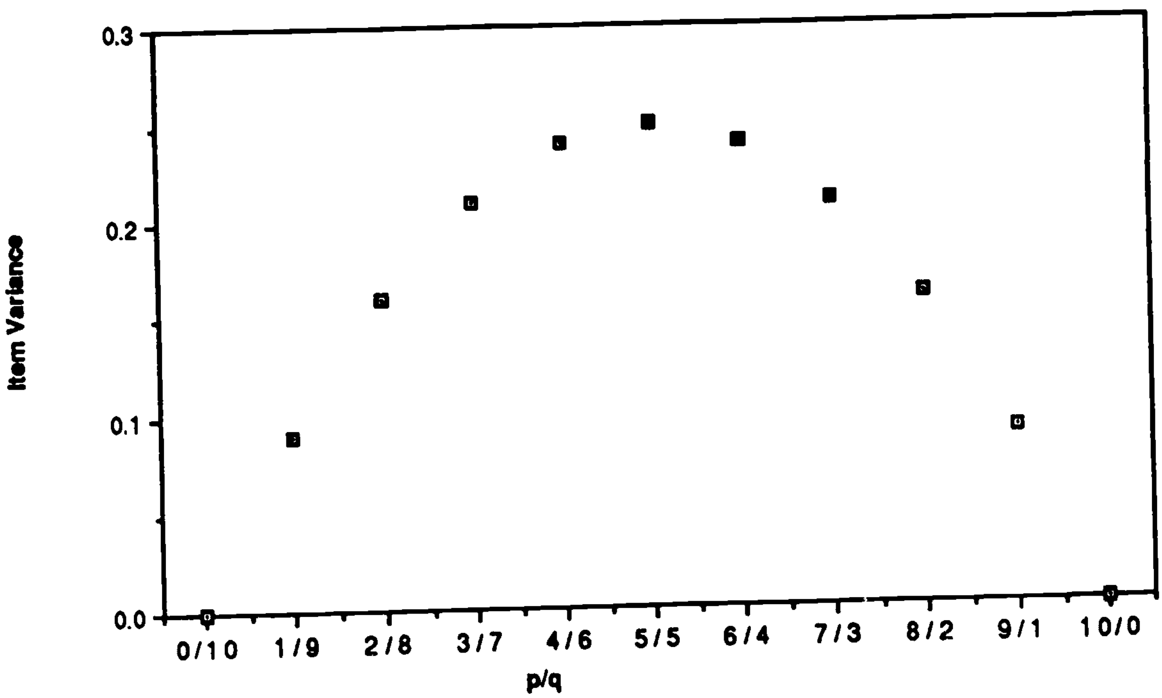

Figure 3
Linear regression line for predicted alpha versus observed alpha